图书情报工作
第 62 卷 第 13 期 2018 年 7 月

基于 Fiddler 代理程序的电子资源使用统计分析系统的设计与应用

■ 陈广
中国科学院福建物质结构研究所 福州 350002

摘要:［目的/意义］对比分析国内图书馆自行采集电子资源使用统计数据的主要方法,解决现有技术方法无法采集基于 HTTPS 协议的电子资源访问信息的问题。［方法/过程］在现有基于旁路监听的采集模式的基础上对技术加以改进,通过策略路由和 Fiddler 代理程序来实现基于 HTTPS 协议访问的电子资源使用数据的统计和分析,并在此基础上设计并应用电子资源使用统计分析系统。［结果/结论］解决了基于 HTTPS 协议访问的电子资源使用数据的采集问题,对其他图书馆自行采集电子资源使用统计数据有一定的借鉴作用。
关键词: 电子资源 策略路由 使用统计数据 Fiddler 代理程序 旁路监听
分类号: G250.7
DOI: 10.13266/j.issn.0252－3116.2018.13.005

1 引言

电子图书、电子期刊、全文数据库、二次文摘数据库等电子资源已经逐渐成为了图书馆用户获取信息的主要途径,其在图书馆资源中所占的比例也越来越重,甚至有部分图书馆已经完全放弃了纸质资源,而将全部经费投入到电子资源的采购当中。电子资源在图书馆中的作用越来越重要,其使用情况的统计数据成为图书馆分析和了解电子资源价值的重要手段[1]。电子资源使用统计数据对于图书馆具有重要价值,其精确地反映了电子资源的利用状况,可为图书馆重构网络门户、提供用户培训课程以及明确重点突出哪些电子资源产品提供重要参考,还能辅助图书馆员制定有关电子资源购买和管理方面的馆藏决策[2]。

数据库商提供的使用统计报告是图书馆获取电子资源使用统计数据的主要方式。大部分数据库商已经可以提供遵守 Counter 使用统计和数据计量标准的统计报告,并支持标准收割接口 SUSHI,可供图书馆统一收集与整合,实现全部电子资源使用统计数据的管理[1]。

基于 Counter 标准的使用统计报告具有规范性、易用性、自动化和低成本的优点,但其也存在一定局限性:①在时效性方面,数据库商的系统基本是在每月的中旬左右生成上个月的使用统计报告。图书馆获取使用统计数据有一定的滞后性。②在全面性方面,Counter 报告只能对订购的电子资源利用情况进行统计,而不能覆盖图书馆的完整网络资源(例如免费的电子资源和馆藏电子资源);Counter 报告只能提供统计数字,无法满足图书馆对电子资源进行内容级/用户级的及时、深入的分析和数据挖掘的需要[3]。

为了解决 Counter 报告的局限性,国内图书馆开展了自行采集电子资源使用统计数据的研究并进行了实际应用。从已有的研究论文来看,图书馆自行采集电子资源使用统计数据根据采用的技术方法主要可以分为 2 种:基于网关日志的采集分析模式和基于旁路监听的采集分析模式。

本文通过研究国内已发表论文在这两种方法上的应用情况来分析这两种方法的优点和局限性。在此基础上,对现有的技术手段上加以改进,设计并实现电子资源使用统计分析系统,并在中国科学院福建物质结构研究所图书情报中心(下文简称"本馆")投入实践使用。

2 图书馆自行采集电子资源使用统计数据的两种方法

2.1 基于网关日志的采集分析模式

用户对 Internet 的访问数据都会经过统一的出口

作者简介:陈广(ORCID:0000－0002－3828－7401),图书情报中心负责人,馆员,E-mail:chenguang@fjirsm.ac.cn。
收稿日期:2018－01－24 修回日期:2018－03－13 本文起止页码:30－36 本文责任编辑:易飞

网关(如核心交换机、防火墙、代理服务器等),出口网关会将所有的访问信息记录在日志上,其中也包含电子资源的访问数据。图书馆通过对日志信息进行采集、过滤和分析就可以得到电子资源的使用统计报告。国内现有的案例有闫晓弟等设计实现电子资源利用统计网关[4],王孝亮等通过防火墙日志挖掘构建电子期刊数据库统计分析系统[5],郭振英等设计电子资源日志统计系统[6],周欣等通过 Web 日志挖掘分析读者行为[7],等等。

通过对上述案例的研究分析可以得知基于网关日志的采集分析模式的优点在于不需要增加额外的网络设备用于抓取分析数据包,而是可以直接利用现有的网关自带的日志功能记录用户访问行为,再通过技术手段过滤、清洗、分析、整合数据后生成电子资源使用统计报告,节省了硬件成本。关键技术的难点在于网关记录的日志数据不仅包含电子资源的访问数据,也包含其他的网络访问信息。在采集电子资源使用统计数据的过程中需要高效、快速、准确地匹配 IP 和特征值数据来过滤掉和用户电子资源访问行为无关的数据。

基于网关日志的采集分析模式也存在一定的局限性:

(1)使用统计报告的唯一数据来源是网关的日志数据。网关记录的日志信息可能不够完整,无法满足图书馆进行深入分析以及数据挖掘的需求。不同网关记录的日志格式区别较大,数据处理相关的关键代码没有通用性。

(2)使用统计报告的生成时间取决于日志数据的收割策略。采用每日收割策略则可以每日生成使用统计报告,如果每月收割一次数据则只能每月生成一次使用统计报告。由于日志数据的特殊性,无法采用实时收割策略,使用统计报告的生成必然有一定的滞后性。滞后性的存在使得基于网关日志的采集分析模式无法实现对电子资源访问行为的实时监控。

2.2 基于旁路监听的采集分析模式

基于旁路监听的采集分析模式是在网络出口的核心设备上通过端口镜像功能复制数据流,将复制的数据流转发至监听分析服务器,由监听分析服务器抓取数据包、解析访问数据、过滤分析后生成电子资源使用统计报告,并在这个过程中对用户的电子资源访问行为进行监控。国内现有的案例有朱玲等对 ERU 和 DRAS 监控与系统的数据获取质量评估探讨[3],张计龙等基于 ERU 系统研究图书馆用户信息行为数据采集方法[8],邹荣利等用旁路监听设计及应用电子资源访问管理与控制系统[9],施晓华等利用旁路监听设计和应用高校电子资源访问管理控制系统[10],王政军等基于旁路监听设计和实现数字资源评估系统[11],武群辉等研究面向科研的高校图书馆电子资源使用统计模式[12]等。

基于旁路监听的采集分析模式有以下几个优点:

(1)直接对用户访问电子资源和电子资源返回的的数据包进行解析,可以得到最原始、准确、完整的电子资源访问信息。

(2)对访问电子资源的数据包的镜像复制和解析是随着用户的信息访问行为实时发生的,也就是说可以实时监控用户的访问行为以及实时生成使用统计报告。及时地监控数据有助于加强电子资源的规范使用,设定违规阈值,对违规行为给予提前预警和处置,避免数据库商大面积封禁而影响其他用户的正常使用[3]。

(3)由于采用了旁路监听模式复制了数据包,不修改原始数据包,不需要改变原有的网络拓扑结构,不会对用户的网络访问行为造成任何影响。

从现有案例的研究分析来看,基于旁路监听的采集分析模式关键的技术难点在于如何抓取数据包、解析数据包提取关键信息。图书馆自行开发电子资源访问监控分析系统的时候,在 Windows 平台下可以采用 WinPcap[10] 或 WireShark 软件抓取解析数据包,在 Linux 平台下则可以使用 NetFilter 框架[11] 或 Iptable 防火墙抓取解析数据包。如果图书馆缺乏技术力量自行开发,也可以考虑采购商业化软件来监控分析电子资源访问信息,如上海光华复旦公司的 ERU 系统和同方知网公司的 DRAS 系统[3]。

相对于基于网关日志的采集分析模式,基于旁路监听的采集分析模式及时性更强,数据更加全面准确,并且不会对网络结构和用户网络访问行为产生影响。旁路监听的技术已经非常成熟,市场上也有商业化的软件,可以说基于旁路监听的采集分析模式是目前最合适图书馆自行采集电子资源使用统计数据的模式。该模式唯一的局限性是需要在原有网络结构上增加专门的监控分析服务器,需要额外的硬件投入,成本较高。

3 现有模式存在的问题及解决方案

3.1 现有模式存在的问题

笔者在对上述的两种模式进行测试的时候发现:采用 HTTP 协议访问的电子资源,两种模式都能采集到电子资源访问信息;而采用 HTTPS 协议访问的电子

资源,无论是日志网关还是监听分析服务器都无法采集到电子资源访问信息,也就是说两种模式对于采用 HTTPS 协议访问的电子资源不起作用。

通过分析发现 HTTP 协议与 HTTPS 协议最大的区别在于:HTTP 协议采用明文传输数据,HTTPS 协议采用密文传输数据。HTTP 协议传输的数据对整个链路上所有的网络设备都是透明的,当数据包到达网关或监听分析服务器时,网关或监听分析服务器可以直接得知传输的内容并加以处理。HTTPS 协议传输的数据只对发送方和接收方透明,传输链路上的其他网络设备都无法得知数据内容[13]。HTTPS 协议的数据加密过程发生在 OSI 七层模型当中的应用层,数据包从网络适配器发送之前就已经加密完成,当数据包到达网关或监听分析服务器时已经是加密后的数据包,网关或监听分析服务器无法对加密的密文解密,仅能抓取到来源 IP 地址、目标 IP 地址和域名 3 个信息,完全无法满足生成电子资源使用统计报告和实时监控用户信息访问行为的需求。

尽管 HTTPS 协议存在访问速度较慢、部署成本高的问题,由于其具有更好的加密性能,可以有效地避免信息泄露,越来越多的网站都从基于 HTTP 协议的访问转向基于 HTTPS 协议的访问。在电子资源方面,国外的 ScienceDirect、Nature、OSA、Springer 数据库都已经采用 HTTPS 访问。随着服务器硬件的发展和技术的进步,HTTPS 的部署成本将越来越低,其访问速度也将得到显著的提升。可以预见将来会有更多的电子资源服务商基于信息安全的考虑转向部署 HTTPS 访问。对图书馆而言,面对越来越多转向部署 HTTPS 访问的电子资源,如何改进现有的技术解决基于 HTTPS 协议访问的用户信息行为无法抓取的问题就成了当务之急。

3.2 基于旁路监听模式的技术改进

解密 HTTPS 协议有几种方法:①利用浏览器存储的私钥解密服务器返回的加密数据;②在应用层加密数据之前,服务器返回的数据被解密之后抓取所需信息;③采用中间人(man-in-the-middle)技术控制客户端和服务端之间的通讯。前两种方法需要在用户 PC 机上安装客户端程序或插件才能实现,对用户影响较大,不适合图书馆采用。第三种方法中的中间人技术,能够与网络通讯两端分别创建连接,交换其收到的数据,使得通讯两端都认为自己直接与对方对话,事实上整个会话都被中间人所控制,用户无法感受到中间人技术对用户信息访问行为的影响[14]。从中间人技术的特点来看,该技术最适合图书馆用于抓取基于 HTTPS 协议访问的用户信息行为。

在旁路监听模式的基础上将支持中间人技术的程序部署在监控分析服务器上,代替原有的数据包捕获程序,用于抓取和分析数据包,就能实现基于 HTTPS 协议访问的用户信息行为的抓取。在支持中间人技术的软件中,Fiddler 软件是一个功能强大的免费的 Web 代理程序,仅需简单配置就能实现解析 HTTPS 协议,记录 HTTP/HTTPS 访问信息,故本文中选取 Fiddler 软件充当中间人程序。

与原有旁路监听模式采用端口镜像技术复制数据包不同的是,中间人技术需要分别同用户及电子资源方建立连接传输数据,这就需要在核心交换机上采用策略路由(policy-based routing,PRB)替代原有的端口镜像技术将用户访问电子资源的数据包转发至中间人程序,中间人程序再根据转发来的数据包信息与通讯双方分别建立连接。策略路由是一种依据用户制定的策略进行路由选择的机制,可以通过策略路由实现数据包的转发。策略路由的应用需要配置路由访问控制列表(access control list,ACL),将电子资源服务器的 IP 地址信息存储在 ACL 列表当中,从而实现数据包的转发和过滤功能。

综上所述,本研究在现有基于旁路监听的采集分析模式的基础上,使用策略路由替代原有的端口镜像技术实现数据包的过滤和转发,在监听分析服务器上使用支持中间人技术的程序替代原有的数据包捕获程序,实现基于 HTTPS 协议访问的用户信息行为的采集,并在此基础上设计和应用电子资源使用统计分析系统。

4 系统功能模块设计与实现

4.1 系统功能模块设计及业务流程

4.1.1 系统功能模块设计 电子资源使用统计分析系统由以下几个功能模块组成:数据过滤转发模块、数据分析模块、统计分析模块和违规监控模块,如图 1 所示:

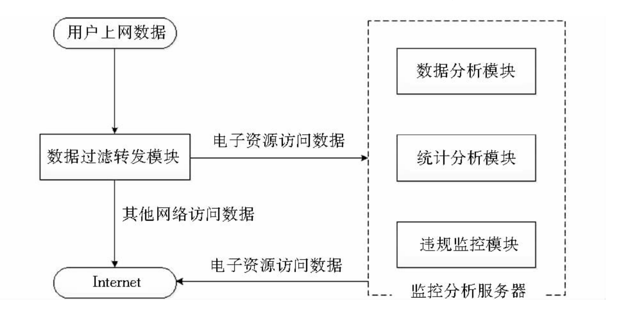


图 1 系统功能模块

数据过滤转发模块在核心交换机上过滤数据包,将电子资源访问相关的数据包转发至监控分析服务器。数据分析模块在监控分析服务器中根据 URL 特征值、HTTP 状态码等信息匹配用户电子资源全文访问行为并记录至 SQL 数据库中。统计分析模块采用 B/S 架构结合特定 SQL 语句展示用户电子资源访问信息。违规监控模块采用 C#语言编写程序监控用户电子资源全文访问次数,对发生违规行为的用户采取警告和封禁两种处理方式。

4.1.2 系统业务流程 当用户产生上网数据时,由数据过滤转发模块过滤掉与电子资源访问无关的数据并将电子资源访问数据转发至数据分析模块;数据分析模块对数据加以分析,记录用户电子资源访问数据当中的全文访问行为,同时触发违规监控模块。违规监控模块更新相应用户的全文下载数,将全文下载数同用户预设的每日预警值和禁用值进行比对,根据比对结果采取对应措施。统计分析模块则独立于业务流程之外,提供电子资源访问数据的可视化等功能。

4.2 系统功能实现

4.2.1 数据过滤转发模块 数据过滤转发功能通过在核心交换机上开启策略路由功能并配置相应的 ACL 列表实现,ACL 列表包含电子资源服务器的 IP 地址信息。数据过滤功能根据数据包的目标 IP 地址进行过滤,当数据包到达核心交换机时,提取数据包的目标 IP 地址,将其与 ACL 列表中存储的电子资源服务器的 IP 地址进行匹配,目标 IP 地址符合的数据包被转发至监控分析服务器,不符合的数据包不作处理。数据转发功能通过指定匹配数据包的下一跳地址为监控分析服务器地址(192.168.4.45)实现。

以 Wiley 数据库为例,其 ACL 匹配代码如下:

```
acl number 3200 description wiley
    rule 1 permit ip destination 199.171.202.195 //Wiley 数据库服务器 IP
if-match acl 3200
    apply ip-address next-hop 192.168.4.45 //监控分析服务器 IP
```

部分电子资源服务器采用 CDN(content delivery network,内容分发网络)加速功能,其 IP 地址的变动性较强,需要动态地维护更新 ACL 列表,以免造成数据包漏发导致用户访问数据的缺失。为了动态更新 ACL 列表,采用 C#语言编写程序,每 10 分钟追踪一次各电子资源服务器当前的 IP 地址,将追踪得到的 IP 信息同 ACL 列表中的 IP 信息进行对比,如果 ACL 列表未包含该 IP 信息则通过核心交换机接口更新 ACL 列表。

4.2.2 数据分析模块 数据分析模块在监控分析服务器上部署 Fiddler 代理程序接收并分析用户电子资源访问数据,通过特征值匹配及一系列条件判断后,将满足判断标准的用户信息行为记录在 SQL 数据库中。

当 Fiddler 程序接收到一次用户访问请求时,首先提取出请求当中目标服务器的域名信息,根据该信息确定对应的 URL 特征值;其次将用户访问请求的 URL 信息与 URL 特征值进行匹配,对于满足特征值信息的请求检测其 HTTP 状态码是否为 200,确定本次请求是否为一次成功的请求;然后对电子资源服务器返回数据的大小进行检测,保障返回的数据内容不为空;最后将符合上述一系列检测的请求进行标记并将相关访问信息记录在 SQL 数据库中。数据分析模块的业务流程如图 2 所示:

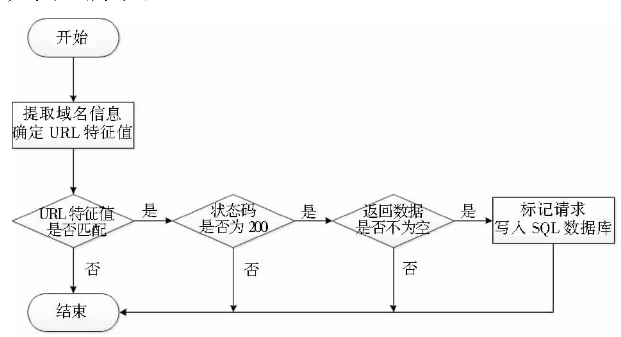


图 2 数据分析模块业务流程

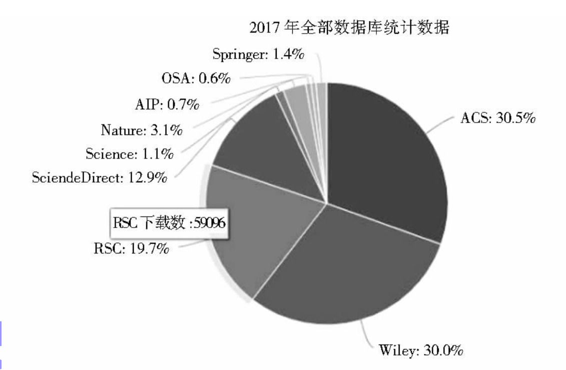
SQL 数据库中为不同的电子资源分别创建对应的数据库,在数据库中以"T + 年份"的形式作为数据表名存储每一年度用户的全文访问信息。数据表中主要包含用户名、部门/课题组、用户 IP、访问 URL、时间等信息,详细的字段设置如表 1 所示:

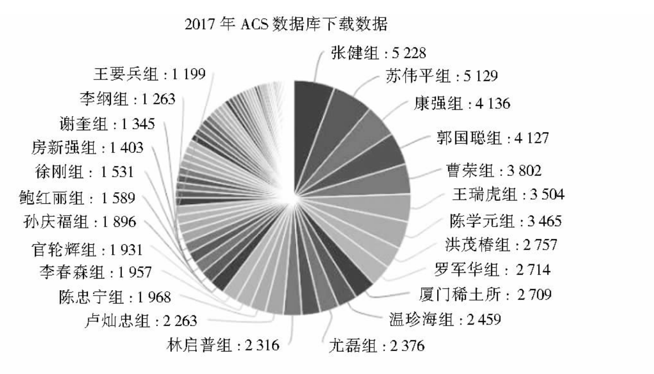
表 1 全文访问数据表

字段名	类型	备注
id	int	唯一标识字段
url	nvarchar(1000)	用户访问的全文 URL
ip	nvarchar(20)	用户 IP 地址
username	nvarchar(100)	用户姓名
year	char(4)	年
month	char(2)	月
day	char(2)	日
hour	char(2)	小时
minute	char(2)	分钟
second	char(2)	秒数
researchgroup	nvarchar(30)	用户所属课题组或部门
type	nvarchar(10)	用户访问电子资源全文类型(PDF 文件或 HTML 文件)

4.2.3 统计分析模块 统计分析模块基于 B/S 架构，采用 PHP 语言，结合 SQL 语句制作网页展示电子资源使用情况，在页面中使用 HighCharts 插件实现电子资源使用统计数据的可视化。

统计分析模块面向图书馆和课题组用户分别提供不同的页面内容。图书馆页面以饼状图及表格形式展示各电子资源全文下载次数和对应比例，同时提供单个电子资源课题组全文总下载次数和对应费用信息（见图 3）。课题组用户页面同样以饼图、柱状图及表格形式展示课题组用户全文下载次数和费用情况，同时提供预警设置页面，用于设置每日预警值和禁用值，避免过量下载（见图 4）。





序号	课题组	下线数	费用
1	张健组	5 228	20 180.08
2	苏伟平组	5 129	19 797.94
3	康强组	4 136	15 964.96

图 3 统计分析模块图书馆页面

预警设置

	预警值	禁用值	
acs:	0	0	修改
wiley:	0	0	修改
rsc:	0	0	修改
sd:	0	0	修改
science:	0	0	修改
nature:	0	0	修改
aip:	0	0	修改
osa:	0	0	修改
springer:	0	0	修改

Email: 请输入电子邮箱地址

PS:课题组数据库全文访问量达到每日预警值时，系统将发送预警邮件至指定的邮箱。达到禁用值时将直接停止课题组在相应数据库的访问权限并同时发送告知邮件。预警值和禁用值应为0或正整数，禁用值应大于等于预警值。如果将数值设置为0，则表示不设置对应预警值或禁用值。

图 4 预警设置页面

4.2.4 违规监控模块 违规监控模块采用 C#语言编写程序，监控数据分析模块中的全文访问数据表。c#程序实时记录每一天各个课题组的全文下载数，当课题组当天的全文下载数达到课题组用户预先设置的每日预警值和禁用值时，分别触发预警和封禁操作。

C#程序中的 sqlDependency 类提供了这样一种功能：当被监测的数据表中的数据发生变化时，会自动触发 OnChange 事件来通知应用程序[15]。在程序中使用 sqlDependency 类监控全文访问数据表，当该表写入了一条新数据时触发 OnChange 事件，程序提取出数据表中新加入记录的 id、researchgroup、ip 等 3 个字段值，根据 researchgroup 字段的值将对应课题组的全文下载数加 1，然后对更新后的课题组全文下载数进行判断。当该数值达到每日预警值时，程序将发送预警邮件至指定的邮箱，达到禁用值时直接停止课题组在相应数据库的访问权限并同时发送告知邮件。每天凌晨程序自动将所有课题组的全文下载数重置为 0。程序运行界面见图 5。

5 系统应用效果和存在的问题

5.1 系统应用效果

本电子资源使用统计分析系统于 2017 年 7 月份部署完成，经多次修改后运行稳定。系统投入使用后，截至 2018 年 1 月 20 日已存储用户全文访问记录 36 万多条，给图书馆提供了强大的数据支持。

在电子资源方面，通过统计分析模块可以了解到各个电子资源的全文下载量、下载量占比、篇均成本等信息，有助于图书馆掌握电子使用情况，调整电子资源保障策略。以 OSA 数据库为例，该数据库 2017 年全文

图 5　违规监控程序界面

下载量低,仅占全部电子资源全文下载量的 0.6%,篇均成本 54.18 元,在本馆全部电子资源中最高。从全文 URL 信息可以得知,该数据库大部分全文下载量来自该数据库免费的 OA 期刊。综合上述信息,可以考虑停止订购该数据库,通过其他方式保障该数据库。

在分析用户信息访问行为方面,通过统计分析模块可以了解到各课题组电子资源使用偏好,有助于图书馆了解科研用户研究方向和科研需求,更好地开展知识化服务工作。另一方面,通过课题组全文下载量数据可以精确计算出课题组所需承担电子资源费用比例,从而向课题组收取部分电子资源费用。

在违规监控方面,系统于 2017 年 9 月份监控到某课题组 ACS 数据库全文下载量异常。通过分析下载记录发现,该课题组使用 EndNote 软件批量下载 ACS 数据库全文。由于系统及时预警了异常下载量,系统管理员临时封禁了该课题组 ACS 数据库的访问权限,避免了因过量下载行为导致数据库商封禁数据库访问权限、影响其他用户正常使用数据库的情况。

5.2　系统存在的问题

系统采用策略路由替代了旁路监听模式中的端口镜像功能,用户的访问数据被直接转发至监控分析服务器,如果监控分析服务器出现故障,无法将用户的访问数据转发至电子资源服务器,就会导致用户无法访问电子资源。要解决这个问题,可以采用备用监控分析服务器的方式,当监控分析服务器故障时,及时启用备用监控分析服务器,保障用户对电子资源的访问,只不过这样又会带来额外的硬件成本。

部分电子资源服务器的 IP 地址处于变动当中。尽管采用了动态更新电子资源 IP 列表的方式,由于更新时间为 10 分钟一次,当电子资源服务器的 IP 地址变更后,有可能 ACL 列表尚未更新。如果这时候有用户访问了该电子资源服务器,则这部分访问数据无法被转发至监控分析服务器,无法被数据分析模块分析记录,从而导致电子资源使用统计数据的不准确。解决这个问题可以考虑通过权限控制方式,只允许从监控分析服务器转发出来的数据包有电子资源访问权限,这样就强制用户访问电子资源的数据包必须通过监控分析服务器才能访问电子资源,从而保障电子资源使用统计数据不会缺失。不过这种方法在电子资源服务器 IP 地址变动而 ACL 列表尚未更新前,会带来短暂的电子资源无法访问的情况。

6　结语

图书馆对电子资源内容级/用户级的及时、深入分析和数据挖掘的需求同数据库商使用统计报告局限性之间的矛盾,促使图书馆需深入研究如何自行采集电子资源使用统计数据。详尽、及时、全面的电子资源使用统计数据对图书馆分析用户需求、避免过量下载行为、提升图书馆知识服务能力有着十分重要的意义。

本文通过对现有旁路监听模式的改进,解决了基于 HTTPS 协议访问的电子资源使用统计数据的采集问题,在此基础上设计并应用了电子资源使用统计分析系统,对其他图书馆自行获取电子资源使用统计数据并加以应用具有一定的借鉴作用。

参考文献:

[1] 陈大庆,叶兰,杨巍,等.电子资源使用统计平台 USSER 的设计与实现[J].图书情报工作,2015,59(1):106-112.
[2] 蔡箐.电子期刊使用数据的统计及标准化研究[J].数字图书馆论坛,2012(7):64-69.
[3] 朱玲,崔海媛.高校图书馆电子资源使用监控与统计系统数据获取质量评估方法探讨[J].图书情报工作,2016,60(5):51-57.
[4] 闫晓弟,邵晶,周奇,等.电子资源利用统计网关系统的设计与实现[J].现代图书情报技术,2018,24(8):97-100.
[5] 王孝亮,王威.通过防火墙日志挖掘构建电子期刊数据库统计分析系统[J].现代图书情报技术,2013,29(z1):122-126.
[6] 郭振英,赵文兵,魏育辉.电子资源日志统计系统分析与设计[J].现代图书情报技术,2008,24(9):102-106.
[7] 周欣,陆康.基于图书馆数字资源访问系统的读者行为数据挖掘研究[J].现代情报,2016,36(1):51-56.
[8] 张计龙,殷沈琴,陈铁.基于 ERU 的图书馆用户信息行为数据采集方法研究——以复旦大学图书馆为例[J].图书馆杂志,2014,33(12):10-16.
[9] 邹荣,张成昱,姜爱蓉,等.电子资源访问管理与控制系统的设计及应用[J].图书情报工作,2010,54(1):121-124.

[10] 施晓华,钱吟,谢锐.高校电子资源访问管理控制系统的设计和应用[J].计算机应用研究,2011,28(3):1042-1045.
[11] 王政军,董晓梅,俞小怡.基于旁路监听的数字资源评估系统的设计与实现[J].图书情报工作,2015,59(9):52-57.
[12] 武群辉,何胜,冯新翎,等.面向科研的高校图书馆电子资源使用统计模式研究[J].新世纪图书馆,2017(11):37-40.
[13] GOURLEY D,TOTTY B. HTTP 权威指南[M].陈涓,赵振平,译.北京:人民邮电出版社,2012.
[14] QU J. 三种解密 HTTPS 流量的方法介绍[EB/OL].[2018-01-22]. https://imququ.com/post/how-to-decrypt-https.html.
[15] SqlDependency 类的使用[EB/OL].[2018-01-22]. https://www.cnblogs.com/lanchong/p/7125400.html.

Design and Application of Electronic Resource Use Statistical Analysis System Based on Fiddler Agent

Chen Guang
Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Fuzhou 350002

Abstract: [**Purpose/significance**] This paper compares and analyzes the main methods of collecting electronic resources using statistical data by libraries in China at present, and solves the problem that the existing technology method cannot acquire the electronic resource access information based on the Https protocol. [**Method/process**] The technology is improved on the basis of the existing collection mode based on bypass monitoring. Through the policy routing and Fiddler agent, the Https protocol based access to electronic resource usage statistics and analysis of data is achieved. Base on this, the electronic resource utilization statistical analysis system is designed and applied. [**Result/conclusion**] This paper solves the problem of collecting data on the usage of electronic resources based on the Https protocol and provides a reference for other libraries to collect their own statistical data on the usage statistics of electronic resources.

Keywords: electronic resource policy-based routing usage statistics data Fiddler proxy bypass monitor